

OPTIMAL ALLOCATION OF OBSERVATIONS FOR PARTITIONING A SET
OF NORMAL POPULATIONS IN COMPARISON WITH A CONTROL^{*}

by

Milton Sobel

Yung Liang Tong^{**}

University of Minnesota

University of Nebraska

Technical Report No. 134

February 1970

University of Minnesota
Minneapolis, Minnesota

^{*}This work was supported by Nebraska Research Council of the University of Nebraska and by National Science Foundation Grant NSF-GP-11021 at the University of Minnesota.

^{**}Work completed while on leave to the University of Minnesota.

1. Introduction.

In many practical situations multiple comparisons of several experimental categories with a standard or control are to be made. For the normal distributions this problem can be treated in the following way: $\pi_0, \pi_1, \dots, \pi_k$ ($k \geq 1$) are $(k + 1)$ normal populations with unknown means μ_i and variances σ_i^2 for $i = 0, 1, \dots, k$, respectively; where π_0 denotes the control population and $\pi_1, \pi_2, \dots, \pi_k$ denote the experimental populations. The experimental populations are compared with the control population simultaneously and multiple decisions about the differences of the population means $\mu_1 - \mu_0, \mu_2 - \mu_0, \dots, \mu_k - \mu_0$ are to be made.

In the literature we have had two different approaches to this multiple decision problem. One approach is the construction of (one-sided and two-sided) rectangular confidence regions for the mean vector

$$(1.1) \quad \underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k) = (\mu_1 - \mu_0, \mu_2 - \mu_0, \dots, \mu_k - \mu_0)$$

based on the multivariate normal or the multivariate generalization of Student's t distribution (see, e.g., Bechhofer [1], Dunnett [3] and Šidák [9]). Another approach is the subset formulation considered by Gupta and Sobel [5] and Lehmann [6]. They discussed the selection of a subset of the k experimental populations which contains all the populations significantly better than the control. Recently Tong [10] considered the solution of this approach under various sampling procedures with vector at a time.

Since every one of the k experimental populations is compared with the control population, it is a natural thing to ask: Are there

advantages in allowing the sample size of the control population to differ from the sample size(s) of the experimental populations? In particular, assuming that all the $(k + 1)$ populations have common variances, if n_0 observations are taken from π_0 and n_1 observations are taken from each of $\pi_1, \pi_2, \dots, \pi_k$, we ask for the "optimal" value of $\alpha = \frac{n_0}{n_1}$ as a function of k when the total sample size n is kept fixed. For the case of one-sided rectangular confidence regions this problem has been considered by Dunnett [3] and, more recently, by Bechhofer [1]. Dunnett considered the situation of unknown but common variances. He prepared several numerical graphs and indicated that the optimal value of α (where the confidence coefficient achieves its maximum) is slightly less than \sqrt{k} for large n . Bechhofer worked on the case of known variances. Under the restriction that the sample means of the experimental populations have common variances, he proved mathematically that the optimal value of α is 0 for small n and it converges to \sqrt{k} as $n \rightarrow \infty$. For moderate n it is the unique root of a certain equation which can be tabulated numerically. It appears that under the subset formulation the corresponding problem of optimal allocation has not been considered yet.

In this paper we assume that the control population has variance σ_0^2 and the experimental populations have a common variance σ_1^2 . We consider the solution of the optimal allocation under the subset formulation for both known and unknown population variances σ_0^2 and σ_1^2 . When the variances are known, it is shown that the expected number of populations misclassified is minimized uniformly in θ and n by taking $\alpha = \sqrt{k\lambda}$ (Theorem 1), where λ is $\lambda = \frac{\sigma_0^2}{\sigma_1^2}$; and the probability of correct

selection at the least favorable configuration is maximized by taking $\alpha = \sqrt{k\lambda}$ when n is large (Theorem 2). Comparison is made between the efficiencies of the present procedure with $\alpha = \sqrt{k\lambda}$ and the procedure with equal number of observations considered in [10]. The relative efficiency of the present procedure is approximately $\frac{2k}{(1 + \sqrt{k})^2}$ for large n and k (Section 4). When the variances are unknown, a sequential sampling procedure which is asymptotically optimal for large n is considered in Section 5. Two related problems of optimal allocation for the construction of two-sided rectangular and elliptical confidence regions are considered and solved by using similar methods in the appendices.

The corresponding solution of this problem when the experimental populations do not have common variances is not considered here. It is hoped that this will be studied in a later paper.

2. Assumptions, Notations and Preliminaries.

Let $\pi_0, \pi_1, \dots, \pi_k$ denote $(k + 1)$ normal populations with unknown means μ_i and variances σ_i^2 for $i = 0, 1, \dots, k$, respectively. Throughout this paper we shall assume that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ and we shall consider only those sampling procedures under which observations from the experimental populations are taken vector at a time. It should be noted that our results hold even under the more general assumption that the sample means of the experimental populations have common variances. However, under the present assumption the notations used are more economical.

For every n_0 and n_1 subject to the condition

$$(2.1) \quad n_0 + kn_1 = n$$

where n is considered fixed, let $\{X_{0j}\}$ ($j = 1, \dots, n_0$) and $\{X_{1j}\}$ ($j = 1, \dots, n_1$) be random samples from π_0 and π_i ($i = 1, 2, \dots, k$) respectively; and let

$$(2.2) \quad \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad i = 0, 1, \dots, k.$$

Denote

$$(2.3) \quad \alpha = \frac{n_0}{n_1},$$

$$(2.4) \quad \lambda = \frac{\sigma_0^2}{\sigma_1^2}$$

and the random vector \underline{Y} by

$$(2.5) \quad \underline{Y} = (Y_1, Y_2, \dots, Y_k) = (\bar{X}_1 - \bar{X}_0, \bar{X}_2 - \bar{X}_0, \dots, \bar{X}_k - \bar{X}_0).$$

Then \underline{Y} has a multivariate normal distribution with mean vector $\underline{\theta}$ defined in (1.1), common variances

$$(2.6) \quad \sigma^2 = \frac{\sigma_1^2}{n} (\alpha + k)(\alpha + \lambda)/\alpha = \frac{\sigma_1^2}{n} g(\alpha) \quad (\text{say}),$$

and common correlation coefficient

$$(2.7) \quad \rho = \frac{\lambda}{\alpha + \lambda}.$$

Note that ρ is monotonically decreasing in α .

We now follow the notations used in [10] to formulate the problem of partitioning the set of k experimental populations. For arbitrary but fixed real numbers δ_1 and δ_2 such that $\delta_1 < \delta_2$, we define three disjoint and exhaustive subsets Ω_B , Ω_I and Ω_G of the set

$$(2.8) \quad \Omega = \{\pi_1, \pi_2, \dots, \pi_k\}$$

by

$$\Omega_B = \{\pi_i: \theta_i \leq \delta_1\},$$

$$(2.9) \quad \Omega_I = \{\pi_i: \delta_1 < \theta_i < \delta_2\},$$

$$\Omega_G = \{\pi_i: \theta_i \geq \delta_2\}.$$

If, for practical purpose, a large population mean indicates advantage, then Ω_B represents the bad subset, Ω_G the good subset and Ω_I the indifference subset. Based on the observed random vector \underline{Y} , we want to partition Ω into two disjoint subsets S_B and S_G such that S_B contains all the significantly bad populations and S_G contains all the significantly good populations. A decision is called a correct decision (CD) if $\Omega_B \subset S_B$ and $\Omega_G \subset S_G$; and a population $\pi_i \in \Omega$ is said to be misclassified if $\pi_i \in (\Omega_B \cap S_G) \cup (\Omega_G \cap S_B)$. Note that a decision is a CD iff the number of populations misclassified is zero.

It follows from an invariant property discussed in [10] that if the multiple decision rule R is such that

$$(2.10) \quad S_B = \{\pi_i: Y_i \leq \frac{1}{2} (\delta_1 + \delta_2)\}$$

$$S_G = \{\pi_i: Y_i > \frac{1}{2} (\delta_1 + \delta_2)\},$$

then the present decision problem depends on $(\mu_0, \mu_1, \dots, \mu_k)$ only through $\underline{\theta}$. For every fixed α we denote

$$(2.11) \quad M_{\underline{\theta}}(\alpha) = \text{expected number of populations misclassified under } R,$$

$$(2.12) \quad \beta_{\underline{\theta}}(\alpha) = P_{\underline{\theta}}[CD|R].$$

If $\Omega_B \cup \Omega_G$ is empty, then $M_{\underline{\theta}}(\alpha) = 0$ and $\beta_{\underline{\theta}}(\alpha) = 1$. Otherwise it is easy to see that

$$(2.13) \quad \begin{aligned} M_{\underline{\theta}}(\alpha) &= \sum_{\{i: \pi_i \in \Omega_B\}} P_{\theta_i} [Y_i > \frac{1}{2} (\delta_1 + \delta_2)] + \sum_{\{i: \pi_i \in \Omega_G\}} P_{\theta_i} [Y_i \leq \frac{1}{2} (\delta_1 + \delta_2)] \\ &= \sum_{\{i: \pi_i \in \Omega_B \cup \Omega_G\}} [1 - \Phi(\frac{\sqrt{n} \zeta_i}{\sigma_1 \sqrt{g(\alpha)}})] \end{aligned}$$

where Φ is the standard normal c.d.f. and

$$(2.14) \quad \zeta_i = |\theta_i - \frac{1}{2} (\delta_1 + \delta_2)|, \quad i = 1, 2, \dots, k.$$

To consider $\beta_{\underline{\theta}}(\alpha)$, it is desirable to look at its infimum over $\underline{\theta}$; because in applications there is no knowledge about $\underline{\theta}$ and we want the probability statement about CD to hold uniformly in $\underline{\theta}$. It follows from a theorem in the Appendix of [10] that for every fixed α let $\underline{\theta}^0$ satisfy

$$(2.15) \quad \inf_{\underline{\theta}} \beta_{\underline{\theta}}(\alpha) = \beta_{\underline{\theta}^0}(\alpha),$$

then $\underline{\theta}^0$ is such that

$$(2.16) \quad \begin{aligned} \theta_1 &= \dots = \theta_m = \delta_1, \\ \theta_{m+1} &= \dots = \theta_k = \delta_2 \end{aligned}$$

with

$$(2.17) \quad m = \begin{cases} k/2 & \text{if } k \text{ is even,} \\ (k+1)/2 & \text{if } k \text{ is odd.} \end{cases}$$

Under this least favorable configuration $\underline{\theta}^0$ we have

$$(2.18) \quad \beta(\alpha) = \beta_{\underline{\theta}^0}(\alpha) = P[Z_i \leq \frac{\sqrt{n} d}{\sigma_1 \sqrt{g(\alpha)}}, i = 1, 2, \dots, k]$$

where $g(\alpha)$ is given in (2.6),

$$(2.19) \quad d = \frac{1}{2} (\delta_2 - \delta_1),$$

and $\underline{Z} = (Z_1, Z_2, \dots, Z_k)$ has a multivariate normal distribution with mean vector $\underline{0}$, variances 1 and correlation matrix (ρ_{ij}) given by

$$(2.20) \quad \rho_{ij} = \begin{cases} \rho & \text{if } i \neq j \text{ and } i, j \in \{1, 2, \dots, m\} \\ & \text{or } i, j \in \{m+1, \dots, k\}, \\ -\rho & \text{if } i \in \{1, 2, \dots, m\} \text{ and } j \in \{m+1, \dots, k\} \end{cases}$$

for ρ specified in (2.7). Applying a standard technique we can write

$$(2.21) \quad \beta(\alpha) = \int_{-\infty}^{\infty} \Phi^m(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}}) \Phi^{k-m}(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}}) \varphi(u) du$$

where φ is the standard normal density and

$$(2.22) \quad b = \frac{d\sqrt{n}}{\sigma_1}.$$

We note from (2.13) and (2.18) that $\beta(\alpha)$ depends on α through both ρ and $g(\alpha)$. However, $M_{\underline{\theta}}(\alpha)$ depends on α only through $g(\alpha)$.

3. The case of known variances.

Throughout this section we shall assume that σ_0^2 , σ_1^2 and λ are known. For every fixed b specified in (2.22) we define α_1 and α_2 to satisfy

$$(3.1) \quad M_{\underline{\theta}}(\alpha_1) = \inf_{\alpha} M_{\underline{\theta}}(\alpha),$$

$$(3.2) \quad \beta(\alpha_2) = \sup_{\alpha} \beta(\alpha).$$

In the following theorem we give an exact solution for α_1 which does not depend on $\underline{\theta}$ and n .

Theorem 1. For every k, n and $\underline{\theta}$, we have

$$(3.3) \quad \alpha_1 = \sqrt{k\lambda}.$$

Proof: From (2.13) it suffices to consider the function

$$g(\alpha) = (\alpha+k)(\alpha+\lambda)/\alpha.$$

Since

$$\frac{d}{d\alpha} \log g(\alpha) = \frac{\alpha^2 - k\lambda}{\alpha(\alpha+k)(\alpha+\lambda)} \begin{matrix} \leq \\ \geq \end{matrix} 0 \text{ for } \alpha \begin{matrix} \leq \\ \geq \end{matrix} \sqrt{k\lambda},$$

it follows that $g(\alpha)$ is monotonically decreasing for $\alpha < \sqrt{k\lambda}$ and monotonically increasing for $\alpha > \sqrt{k\lambda}$. Therefore $g(\alpha)$ achieves its minimum at $\alpha = \sqrt{k\lambda}$ and the proof of the theorem is completed.

Theorem 1 asserts that if the goal of the optimal allocation of observations is to minimize the expected number of populations misclassified, then this goal is achieved, uniformly in $\underline{\theta}$ and n , by taking $\alpha = \sqrt{k\lambda}$. In particular if all the $(k+1)$ populations have common variances ($\lambda=1$), then the number of observations from the control population should be \sqrt{k} times of the number of observations from each one of the k experimental populations.

We now proceed to find the solution of α_2 which maximizes the probability of CD under the least favorable configuration $\underline{\theta}^0$. Since differentiation under the integral sign is permissible here, α_2 must satisfy

$$(3.4) \quad \beta'(\alpha_2) = 0$$

where

$$(3.5) \quad \beta'(\alpha) = \int_{-\infty}^{\infty} \left[\frac{d}{d\alpha} \Phi^m \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m} \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \right] \varphi(u) du.$$

If $k = 1$, the well-known result is $\alpha_2 = \sqrt{\lambda}$. In the following we shall consider only $k \geq 2$ (i.e., $m \geq 1$ and $(k-m) \geq 1$).

After integrating by parts it is easy to see that

$$\begin{aligned} \Delta_1 &= - \int_{-\infty}^{\infty} \Phi^{m-1} \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m} \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) u \varphi(u) du \\ &= \frac{\sqrt{\alpha\lambda}}{(\alpha+\lambda)} \left[\frac{b}{\sqrt{\alpha+k}} A_1 + (k-m)A_3 - (m-1)A_4 \right], \\ \Delta_2 &= - \int_{-\infty}^{\infty} \Phi^m \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m-1} \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) u \varphi(u) du \\ &= \frac{\sqrt{\alpha\lambda}}{(\alpha+\lambda)} \left[- \frac{b}{\sqrt{\alpha+k}} A_2 - mA_3 + (k-m-1)A_5 \right]; \end{aligned}$$

where $A_1 - A_5$ are the following multivariate normal probability integrals

$$(3.6) \quad A_1 = \int_{-\infty}^{\infty} \Phi^{m-1} \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m} \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi(u) du,$$

$$(3.7) \quad A_2 = \int_{-\infty}^{\infty} \Phi^m \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m-1} \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi(u) du,$$

$$(3.8) \quad A_3 = \int_{-\infty}^{\infty} \Phi^{m-1} \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m-1} \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \\ \cdot \varphi \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi(u) du,$$

$$(3.9) \quad A_4 = \int_{-\infty}^{\infty} \Phi^{m-2} \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m} \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi^2 \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi(u) du,$$

$$(3.10) \quad A_5 = \int_{-\infty}^{\infty} \Phi^m \left(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m-2} \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi^2 \left(-\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}} \right) \varphi(u) du.$$

Since by (3.5)

$$\beta'(\alpha) = \frac{1}{2} \left[\sqrt{\frac{\lambda}{\alpha^3}} \{m\Delta_1 - (k-m)\Delta_2\} - \frac{b}{\sqrt{(\alpha+k)^3}} \{mA_1 + (k-m)A_2\} \right],$$

it follows that $\beta'(\alpha) = 0$ iff

$$(3.11) \quad \frac{(k\lambda - \alpha^2)b}{\lambda\sqrt{(\alpha+k)^3}} [mA_1 + (k-m)A_2] = m(m-1)A_4 + (k-m)(k-m-1)A_5 - 2m(k-m)A_3.$$

Now if we complete the squares in $A_1 - A_5$ and change variables in the integrals, (3.11) will have the form

$$\begin{aligned}
(3.12) \quad & \frac{(k\lambda - \alpha^2)b}{\lambda\sqrt{(\alpha+k)^3}} \sqrt{\frac{\alpha}{\alpha+\lambda}} \varphi\left(\frac{b\sqrt{\alpha}}{\sqrt{(\alpha+\lambda)(\alpha+k)}}\right) [mI_1 + (k-m)I_2] \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\alpha}{\alpha+2\lambda}} \left[\varphi\left(\frac{b\sqrt{2\alpha}}{\sqrt{(\alpha+2\lambda)(\alpha+k)}}\right) \{m(m-1)I_4 + (k-m)(k-m-1)I_5\} \right. \\
&\quad \left. - 2m(k-m)\varphi\left(\frac{b\sqrt{2}}{\sqrt{\alpha+k}}\right)I_3 \right]
\end{aligned}$$

where

$$(3.13) \quad I_1 = \int_{-\infty}^{\infty} \Phi^{m-1} \left(\sqrt{\frac{\lambda}{\alpha+\lambda}} v + \frac{\alpha b}{(\alpha+\lambda)\sqrt{\alpha+k}} \right) \Phi^{k-m} \left(-\sqrt{\frac{\lambda}{\alpha+\lambda}} v + \frac{(\alpha+2\lambda)b}{(\alpha+\lambda)\sqrt{\alpha+k}} \right) \varphi(v) dv,$$

$$(3.14) \quad I_2 = \int_{-\infty}^{\infty} \Phi^m \left(\sqrt{\frac{\lambda}{\alpha+\lambda}} v + \frac{(\alpha+2\lambda)b}{(\alpha+\lambda)\sqrt{\alpha+k}} \right) \Phi^{k-m-1} \left(-\sqrt{\frac{\lambda}{\alpha+\lambda}} v + \frac{\alpha b}{(\alpha+\lambda)\sqrt{\alpha+k}} \right) \varphi(v) dv,$$

$$(3.15) \quad I_3 = \int_{-\infty}^{\infty} \Phi^{m-1} \left(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{b}{\sqrt{\alpha+k}} \right) \Phi^{k-m-1} \left(-\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{b}{\sqrt{\alpha+k}} \right) \varphi(v) dv,$$

$$(3.16) \quad I_4 = \int_{-\infty}^{\infty} \Phi^{m-2} \left(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{\alpha b}{(\alpha+2\lambda)\sqrt{\alpha+k}} \right) \Phi^{k-m} \left(-\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{(\alpha+4\lambda)b}{(\alpha+2\lambda)\sqrt{\alpha+k}} \right) \varphi(v) dv,$$

$$(3.17) \quad I_5 = \int_{-\infty}^{\infty} \Phi^m \left(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{(\alpha+4\lambda)b}{(\alpha+2\lambda)\sqrt{\alpha+k}} \right) \Phi^{k-m-2} \left(-\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{\alpha b}{(\alpha+2\lambda)\sqrt{\alpha+k}} \right) \varphi(v) dv$$

are one-sided multivariate normal probabilities whose covariance matrix have the form described in (2.20), and they approach to 1 as $b \rightarrow \infty$. We note that $I_3 = 1$ and $I_4 = I_5 = 0$ for $k = 2$. Define

$$(3.18) \quad x = \frac{b}{\sqrt{(\alpha+\lambda)(\alpha+2\lambda)(\alpha+k)}}.$$

Since

$$\frac{1}{\sqrt{2\pi}} \varphi\left(\frac{b\sqrt{2\alpha}}{\sqrt{(\alpha+2\lambda)(\alpha+k)}}\right) / \varphi\left(\frac{b\sqrt{\alpha}}{\sqrt{(\alpha+\lambda)(\alpha+k)}}\right) = \varphi(\alpha x),$$

$$\frac{1}{\sqrt{2\pi}} \varphi\left(\frac{b\sqrt{2}}{\sqrt{\alpha+k}}\right) / \varphi\left(\frac{b\sqrt{\alpha}}{\sqrt{(\alpha+\lambda)(\alpha+k)}}\right) = \varphi((\alpha+2\lambda)x),$$

finally (3.12) can be written as

$$(3.19) \quad c_1(k\lambda - \alpha^2) = c_2 \frac{\varphi(\alpha x)}{x} - c_3 \frac{\varphi((\alpha+2\lambda)x)}{x},$$

where

$$(3.20) \quad c_1 = (\alpha+2\lambda)[mI_1 + (k-m)I_2] / [\lambda(\alpha+k)],$$

$$(3.21) \quad c_2 = m(m-1)I_4 + (k-m)(k-m-1)I_5, \text{ and}$$

$$(3.22) \quad c_3 = 2m(k-m)I_3.$$

Thus far we have shown that the optimal value α_2 must satisfy (3.19). We summarize the behavior of α_2 in the following theorem.
Theorem 2. Let α_2 be defined in (3.2).

- (a) For $k = 2$, $\alpha_2 > \sqrt{k\lambda}$ for every b and every λ ;
- (b) for $k > 2$, $\alpha_2 < \sqrt{k\lambda}$ for every b and large λ ;
- (c) for every $k \geq 2$ and every λ , $\alpha_2 \rightarrow \sqrt{k\lambda}$ as $b \rightarrow \infty$.

Proof: We first note that $C_i \geq 0$ for $i = 1, 2, 3$. If $k = 2$, then $C_2 = 0$, $C_3 = 1$ and the r.h.s. of (3.19) is negative. In order for the equality to hold we must have $k\lambda < \alpha^2$. This proves (a). The proof of (b) follows from a similar argument. (c) follows from the fact that as $b \rightarrow \infty$ ($x \rightarrow \infty$), the r.h.s. of (3.19) approaches to zero. This completes the proof of the theorem.

If $k = 2$, then $I_1 = I_2$ is a univariate normal probability and (3.19) reduces to

$$(3.23) \quad \frac{\varphi(y)}{y} = \frac{(\alpha^2 - 2\lambda)}{\lambda(\alpha + 2)} \Phi(y)$$

with

$$(3.24) \quad y = (\alpha + 2\lambda)x = \frac{b\sqrt{(\alpha + 2\lambda)}}{\sqrt{(\alpha + \lambda)(\alpha + 2)}}.$$

Using the well-known result that $\frac{\varphi(y)}{y} \sim 1 - \Phi(y)$ for large y , it follows that for $k = 2$ α_2 can be approximated by the solution of the equation

$$(3.25) \quad \Phi(y) = \frac{\lambda\alpha + 2\lambda}{\lambda\alpha + \alpha^2}$$

for large b .

Equation (3.25) suggests (at least for $k = 2$) that, instead of solving (3.19) in terms of α for fixed b , for practical purpose it is more convenient to consider the solution in terms of b for fixed α and then look at its inverse. An algorithm for this purpose is given with examples for $k = 2$ in the next section of this paper. For $k > 2$, in general we will need new multivariate normal tables to evaluate

the integrals $I_1 - I_5$. However, if k is an even number ($m = \frac{k}{2}$), then we have $I_1 = I_2$ and $I_4 = I_5$. Hence for even $k > 2$ there are only three integrals involved in the solution of (3.19).

We observe the significance of the asymptotic result given in Theorem 2. Since for every $k \geq 2$ and $\lambda > 0$ we have $\alpha_2 \rightarrow \sqrt{k\lambda}$ as $b \rightarrow \infty$, it follows that if the total sample size n is large, for practical purposes we can take α to be $\sqrt{k\lambda}$. Combining this with Theorem 1 we can say that if n is large, taking the optimal value of α to be $\sqrt{k\lambda}$ will achieve both the minimization of the expected number of populations misclassified and the maximization of the probability of correct decision at the same time.

4. Illustration and comparison.

In this section we first develop a fairly rapidly converging algorithm for the exact value of α_2 and illustrate its application for the case of $k = 2$. Then we shall compare the efficiency of the present procedure for $\alpha = \sqrt{k\lambda}$ with the procedure of equal number of observations considered in [10].

As mentioned in Section 3, for practical purposes it is more convenient to solve (3.19) in terms of b for fixed α and then look at its inverse. When $k = 2$, b and α_2 are related by (3.23) where y is a linear function of b given in (3.24). Hence for fixed α we want to solve (3.23) in terms of y . At the first step if we take $\Phi(y)$ to be one and solve y_0 from

$$\frac{\varphi(y_0)}{y_0} = \frac{(\alpha^2 - 2\lambda)}{\lambda(\alpha + 2)},$$

then y_0 serves as the first approximation to y . Now we substitute y_0 in (3.23) and solve

$$\frac{\varphi(y_1)}{y_1} = \frac{(\alpha^2 - 2\lambda)}{\lambda(\alpha + 2)} \varphi(y_0)$$

for y_1 ; this y_1 is a more accurate approximation to y . If the same process is repeated over and over, then $\{y_j\}$ will converge to y rapidly, and from this y value we can compute the corresponding b value from (3.24). Although it has not been proved, it has been noted that the y_j -values alternately form upper and lower bounds for the true value of y as they converge to b ; this helps to determine the accuracy of the final result. For the purpose of illustration, we consider two examples.

- (a) Example 1. $k = 2$, $\lambda = 1$, $\alpha = 3$. $y_0 = 1.095$, $y_1 = 1.168$, $y_2 = 1.159$, $y_3 = 1.160$, $y_4 = 1.160 = y$, $b = 2.3200$.
- (b) Example 2. $k = 2$, $\lambda = 1$, $\alpha = 1.45$ (note that α is very close to $\sqrt{k\lambda} = \sqrt{2}$). $y_0 = 1.961$, $y_1 = 1.972$, $y_2 = 1.972 = y$, $b = 3.0867$. The approximated b value computed from (3.25) is 2.9708.

It appears that for most α and λ values the convergence is fairly rapid. Therefore, for $k = 2$, we do not include here the table of the α values, since the process illustrated above can be carried out quite easily.

For $k > 2$, the same algorithm can be used based on (3.19). In this case the multivariate normal probabilities have not been tabulated yet. Tables for this purpose are being computed and they are expected to be available soon.

We now compare the efficiencies of the present procedure and the procedure of equal number of observations. We shall consider here only the case $\lambda = 1$. For fixed P^* , the probability of correct decision at the least favorable configuration $\underline{\theta}^0$, let N' be the total sample size required to satisfy the probability requirement under the present procedure with $\alpha = \sqrt{k\lambda} = \sqrt{k}$, and N'' be the sample size required when the number of observations are equal for all $(k+1)$ populations. We note that N' is greater than the actual sample size required under the optimal allocation with $\alpha = \alpha_2$, but their ratio should be approximately 1 when k and P^* are large. From Theorem 1.1 of [10] we have

$$(4.1) \quad N'' = 2k\left(\frac{\sigma_0}{d}\right)^2 b_0^2$$

where σ_0^2 is the common population variance and the values of b_0 are tabulated in Table 1 of [10]. To find the exact expression of N'' we would need new tables; however, following from (2.18) and a similar argument given in the proof of Theorem 1.2 of [10] we have

$$(4.2) \quad N' < g(\alpha)\left(\frac{\sigma_0}{d}\right)^2 b'^2 = (1 + \sqrt{k})^2 \left(\frac{\sigma_0}{d}\right)^2 b'^2$$

where b' can be computed from multivariate normal tables with equal correlation coefficients. When k is an even number, b' is the solution of the equation

$$(4.3) \quad P[Z_i \leq b', i = 1, \dots, \frac{k}{2}] = \frac{1}{2}(1 + P^*)$$

where $(Z_1, \dots, Z_{k/2})$ has a multivariate normal distribution with mean vector $\underline{0}$, variances 1 and common correlation coefficient $\rho = 1/(1+\sqrt{k})$. Those b' values can be found from tables of Gupta [4]. Combining (4.1)

and (4.2) we have, for every d, σ_0^2, k and P^* ,

$$(4.4) \quad \frac{N''}{N'} > \frac{2k}{(1+\sqrt{k})^2} \left(\frac{b_0}{b'}\right)^2.$$

The ratio on the r.h.s. of (4.4) represents the "conservative" measurement of the relative efficiency of the present procedure for $\alpha = \sqrt{k\lambda}$ w.r.t. the procedure of equal number of observations. For example, if $k = 16$ and $P^* = 0.95$, we have, from Table 1 of [10] and [4], $b_0 = 2.6521756$, $b' = 2.71905$, $\left(\frac{b_0}{b'}\right)^2 = 0.951415$ and the r.h.s. of (4.4) is 1.2178. It appears that $\left(\frac{b_0}{b'}\right) \rightarrow 1$ as k and P^* (or n) become large. In this case the r.h.s. of (4.4) is approximately $2k/(1+\sqrt{k})^2$, which is monotonically increasing in k and it approaches to 2 as $k \rightarrow \infty$. This shows that when the number of experimental populations k is very large, the advantage of the present procedure becomes more significant.

5. An asymptotically optimal procedure for the case of unknown variances.

In this section we shall consider the case that the variances σ_0^2 , σ_1^2 and λ are completely unknown. We shall consider a sequential sampling procedure for the purpose of optimal allocation when the total sample size n is fixed. This sequential procedure has the nature that, at each stage, a decision is made on whether we should take one observation from the control or take a vector of observations from the k experimental populations. Using the results in [7] we can show that under this sequential sampling procedure the allocation of observations is asymptotically optimal when the fixed total sample size n is large.

If σ_0^2 , σ_1^2 and λ were known and if we take $\alpha = \frac{n_0}{n_1}$ to be $\sqrt{k\lambda}$, then (from Section 3) the expected number of populations misclassified

will be minimized uniformly in $\underline{\theta}$ for every n , and the probability of correct decision at $\underline{\theta}^0$ will be maximized for large n . Therefore we should consider a sequential procedure so that when n is large, $\frac{N_0}{N_1}$ will converge to $\sqrt{k\lambda}$ (in some sense), where N_0 is the random sample size from the control and N_1 is the random sample size from each one of the k experimental populations. Motivating from the sequential asymptotic solution to the Behrens-Fisher problem given in [7], we consider the following

Sequential Procedure:

(a) Let $n^* \geq 2$ be a predetermined positive integer. We first take n^* observations from each one of the $(k+1)$ populations.

(b) At each stage after observing r observations from π_0 and s observations from each one of $\pi_1, \pi_2, \dots, \pi_k$, we compute the sample standard deviations

$$(5.1) \quad u_r = \left[\frac{1}{r-1} \sum_{j=1}^r \{x_{0j} - (\frac{1}{r} \sum_{j=1}^r x_{0j})\}^2 \right]^{1/2},$$

$$(5.2) \quad v_s = \left[\frac{1}{k(s-1)} \sum_{i=1}^k \sum_{j=1}^s \{x_{ij} - (\frac{1}{s} \sum_{j=1}^s x_{ij})\}^2 \right]^{1/2}.$$

(c) For the next stage we take one observation from π_0 if $\frac{r}{s} \leq \frac{\sqrt{k} u_r}{v_s}$, or we take a vector of observations from $\pi_1, \pi_2, \dots, \pi_k$ if $\frac{r}{s} > \frac{\sqrt{k} u_r}{v_s}$.

(d) Stop sampling when $r + ks = n$ and apply the decision rule R specified in (2.10).

Let N_0 and N_1 denote the random sample sizes satisfying $N_0 + kN_1 = n$ (fixed). By the strong law of large numbers we have $\frac{\sqrt{k} u_r}{v_s} \rightarrow \sqrt{k\lambda}$ as $r, s \rightarrow \infty$. Considering the distribution of the random variable $\frac{N_0}{N_1}$, under this sequential sampling rule the expected number of populations misclassified is $EM_{\underline{\theta}}(\frac{N_0}{N_1})$ and the probability of correct decision at $\underline{\theta}^0$ is $E\beta(\frac{N_0}{N_1})$, where the functions M and β are defined in (2.13) and (2.21), respectively. We summarize our results in the following theorem. Since the proof of this theorem follows from a similar argument given in [7], it is omitted here.

Theorem 3. Under the above-defined sequential procedure for every n^* we have, as $n \rightarrow \infty$,

$$(5.3) \quad \frac{N_0}{N_1} \rightarrow \sqrt{k\lambda} \text{ a.s.},$$

$$(5.4) \quad EM_{\underline{\theta}}(\frac{N_0}{N_1}) \rightarrow M_{\underline{\theta}}(\sqrt{k\lambda}),$$

$$(5.5) \quad E\beta(\frac{N_0}{N_1}) \rightarrow \beta(\sqrt{k\lambda}).$$

Theorem 3 asserts that if σ_0^2 , σ_1^2 and λ are unknown and if the n observations are allocated according to the above sequential procedure, then the allocation is asymptotically optimal when n is large.

Appendix A

Optimal Allocation of Observations for Two-Sided

Rectangular Confidence Regions

In this appendix we shall employ the same technique used in Section 3 of this paper to investigate the related problem of optimal allocation of observations for the construction of two-sided rectangular confidence regions; and we shall follow the same notations developed in Section 2. For arbitrary but fixed constant vector

$$(A.1) \quad \underline{d} = (d_1, d_2, \dots, d_k), d_i > 0$$

let the two-sided confidence region R_1 be defined by

$$(A.2) \quad R_1 = \{\underline{\omega} = (\omega_1, \omega_2, \dots, \omega_k) : |Y_i - \omega_i| \leq d_i, i = 1, 2, \dots, k\};$$

where $\underline{Y} = (Y_1, Y_2, \dots, Y_k)$ is given in (2.5). It is easy to see that the confidence coefficient for R_1 is

$$(A.3) \quad \gamma(\alpha) = P_{\underline{\theta}}[\underline{\theta} \in R_1] = P_{\rho(\alpha)}[|Z_i| \leq \frac{b_i}{\sqrt{g(\alpha)}}, i = 1, 2, \dots, k]$$

where $b_i = \frac{d_i \sqrt{n}}{\sigma_1}$ for $i = 1, 2, \dots, k$; and $\underline{Z} = (Z_1, Z_2, \dots, Z_k)$ has a multivariate normal distribution with mean vector $\underline{0}$, variances 1 and common correlation coefficient $\rho = \rho(\alpha)$ given in (2.7). We define α_3 to satisfy

$$(A.4) \quad \gamma(\alpha_3) = \sup_{\alpha} \gamma(\alpha)$$

and we shall investigate the behavior of α_3 as a function of k and n .

If $k = 1$, the well-known result is $\alpha_3 = \sqrt{\lambda}$ which does not depend on n . For $k \geq 2$ if in the special case $d_1 = d_2 = \dots = d_k = d$, then applying the standard technique we can write

$$(A.5) \quad \gamma(\alpha) = \int_{-\infty}^{\infty} [\Phi(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}}) - \Phi(\sqrt{\frac{\lambda}{\alpha}} u - \frac{b}{\sqrt{\alpha+k}})]^k \varphi(u) du$$

with $b = \frac{d\sqrt{n}}{\sigma_1}$. Hence if α_3 satisfy (A.4), it must satisfy

$$(A.6) \quad \int_{-\infty}^{\infty} \frac{d}{d\alpha} [\Phi(\sqrt{\frac{\lambda}{\alpha}} u + \frac{b}{\sqrt{\alpha+k}}) - \Phi(\sqrt{\frac{\lambda}{\alpha}} u - \frac{b}{\sqrt{\alpha+k}})]^k \varphi(u) du = 0.$$

After carrying out similar computations given in Section 3, (A.6) can be written as

$$(A.7) \quad D(k\lambda - \alpha^2) = (I'_4 + I'_5) \frac{\varphi(\alpha x)}{x} - 2I'_3 \frac{\varphi((\alpha+2\lambda)x)}{x}$$

where x is given in (3.18),

$$(A.8) \quad D = (\alpha+2\lambda)(I'_1 + I'_2) / [\lambda(k-1)(\alpha+k)]$$

and the two-sided multivariate normal probability integrals $I'_1 - I'_5$ are

$$(A.9) \quad I'_1 = \int_{-\infty}^{\infty} [\Phi(\sqrt{\frac{\lambda}{\alpha+\lambda}} v + \frac{\alpha b}{(\alpha+\lambda)\sqrt{\alpha+k}}) - \Phi(\sqrt{\frac{\lambda}{\alpha+\lambda}} v - \frac{(\alpha+2\lambda)b}{(\alpha+\lambda)\sqrt{\alpha+k}})]^{k-1} \varphi(u) du,$$

$$(A.10) \quad I'_2 = \int_{-\infty}^{\infty} [\Phi(\sqrt{\frac{\lambda}{\alpha+\lambda}} v + \frac{(\alpha+2\lambda)b}{(\alpha+\lambda)\sqrt{\alpha+k}}) - \Phi(\sqrt{\frac{\lambda}{\alpha+\lambda}} v - \frac{\alpha b}{(\alpha+\lambda)\sqrt{\alpha+k}})]^{k-1} \varphi(u) du,$$

$$(A.11) \quad I'_3 = \int_{-\infty}^{\infty} [\Phi(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{b}{\sqrt{\alpha+k}}) - \Phi(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v - \frac{b}{\sqrt{\alpha+k}})]^{k-2} \varphi(u) du,$$

$$(A.12) \quad I_4' = \int_{-\infty}^{\infty} \left[\Phi\left(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{\alpha b}{(\alpha+2\lambda)\sqrt{\alpha+k}}\right) - \Phi\left(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v - \frac{(\alpha+4\lambda)b}{(\alpha+2\lambda)\sqrt{\alpha+k}}\right) \right]^{k-2} \varphi(v) dv,$$

$$(A.13) \quad I_5' = \int_{-\infty}^{\infty} \left[\Phi\left(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v + \frac{(\alpha+4\lambda)b}{(\alpha+2\lambda)\sqrt{\alpha+k}}\right) - \Phi\left(\sqrt{\frac{\lambda}{\alpha+2\lambda}} v - \frac{\alpha b}{(\alpha+2\lambda)\sqrt{\alpha+k}}\right) \right]^{k-2} \varphi(v) dv.$$

Those integrals approach to 1 as $b \rightarrow \infty$. For $k = 2$, $I_3' = I_4' = I_5' = 1$ and I_1' , I_2' are univariate standard normal probabilities. In this special case (A.7) reduces to

$$(A.14) \quad \frac{\varphi(\alpha x)}{x} - \frac{\varphi((\alpha+2\lambda)x)}{x} = \frac{(\alpha+2\lambda)(2\lambda-\alpha^2)}{\lambda(\alpha+2)} [\Phi(\alpha x) + \Phi((\alpha+2\lambda)x) - 1].$$

The numerical values of α_3 can then be tabulated based on (A.7) or (A.14) depending on $k > 2$ or $k = 2$. In the following theorem we observe the general behavior of α_3 .

Theorem A. Let α_3 satisfy (A.4).

(a) For every k and every (d_1, d_2, \dots, d_k) ,

$$(A.15) \quad \alpha_3 \leq \sqrt{k\lambda} \quad \text{for every } \mathbf{d}.$$

(b) For every k ,

$$(A.16) \quad \alpha_3 \rightarrow \sqrt{k\lambda} \quad \text{as } \mathbf{d} \rightarrow \infty$$

when $d_1 = d_2 = \dots = d_k$.

Proof: To prove (a) we use (A.3). Since (from the proof of Theorem 1)

$g(\alpha)$ is monotonically increasing for $\alpha > \sqrt{k\lambda}$ and $\rho = \rho(\lambda)$ is

monotonically decreasing for all α , it follows that for every $\alpha' > \sqrt{k\lambda}$

we have

$$\gamma(\alpha') = P_{\rho(\alpha')} [|Z_i| \leq \frac{b_i}{\sqrt{g(\alpha')}}], i = 1, 2, \dots, k]$$

$$< P_{\rho(\alpha')} [|Z_i| \leq \frac{b_i}{\sqrt{g(\sqrt{k\lambda})}}], i = 1, 2, \dots, k]$$

$$\leq P_{\rho(\sqrt{k\lambda})} [|Z_i| \leq \frac{b_i}{\sqrt{g(\sqrt{k\lambda})}}], i = 1, 2, \dots, k] = \gamma(\sqrt{k\lambda}) ;$$

where the second inequality follows from a recent result of Šidák [9]. This proves (a). If $d_1 = d_2 = \dots = d_k = d$, we examine the behavior of α which satisfies (A.7) for large b . Clearly as $b \rightarrow \infty$ ($x \rightarrow \infty$) the r.h.s. of (A.7) approaches to 0. In order for the equality to hold we must have $\alpha \rightarrow \sqrt{k\lambda}$. This completes the proof of the theorem.

We note that for the construction of the one-sided rectangular confidence regions an inequality similar to (A.15) can also be obtained. The only difference for the one-sided case is that we use the well-known inequality of Slepian instead of Šidák's inequality in the proof. This result has been obtained by Bechhofer in his paper [1:p. 468] by using direct computations for the special case of $d_1 = d_2 = \dots = d_k$. However, our result is more general because we do not require the d_i 's to be equal.

We also note that if the variances σ_0^2 , σ_1^2 and λ are unknown, then a corresponding sequential procedure similar to that used in Section 5 can be carried out for the construction of rectangular confidence regions. The results in Section 5 will also apply to this case and, as $n \rightarrow \infty$, the procedure is asymptotically optimal.

Appendix B

Optimal Allocation of Observations for Elliptical Confidence Regions

In this appendix we consider the problem of optimal allocation of observations for the construction of elliptical confidence regions. For every α the random vector $\underline{Y} = (Y_1, Y_2, \dots, Y_k)$ defined in (2.5) has a multivariate normal distribution with mean vector $\underline{0}$, variances σ^2 given in (2.6) and a common correlation coefficient ρ given in (2.7). Let the $(k \times k)$ correlation matrix \underline{Z} be

$$(B.1) \quad \underline{Z} = \begin{cases} 1 & \text{for } i = j \\ \rho & \text{for } i \neq j \end{cases}$$

and the elliptical region for $\underline{\theta}$ be

$$(B.2) \quad R_2 = \{\underline{w} = (w_1, w_2, \dots, w_k) : (\underline{Y} - \underline{w})(\sigma^2 \underline{Z})^{-1}(\underline{Y} - \underline{w})' \leq c\}$$

with $c > 0$. Then the confidence coefficient for R_2 is

$$(B.3) \quad P_{\underline{\theta}}[\underline{\theta} \in R_2] = P[\chi_k^2 \leq c]$$

which can be found from the chi-square distributions. Since the determinant of \underline{Z} is $|\underline{Z}| = (1-\rho)^{k-1}[(k-1)\rho + 1]$, it follows (see, e.g., [2:p. 120]) that for every fixed k, λ, α and c the volume of R_2 is

$$\begin{aligned} V(\alpha, c) &= [\Gamma(\frac{k}{2} + 1)]^{-1} \pi^{k/2} c^{k/2} \sqrt{|\sigma^2 \underline{Z}|} \\ &= M[(\alpha+k)^k (\alpha+k\lambda)/\alpha]^{1/2} c^{k/2} \end{aligned}$$

where

$$M = [\Gamma(\frac{k}{2} + 1)]^{-1} (\pi \sigma_1^2 / n)^{k/2}$$

does not depend on α and c . If the volume of R_2 is kept fixed at V_0 , α and c are related by

$$(B.4) \quad c = c(\alpha) = \left(\frac{V_0}{M}\right)^{2/k} (\alpha+k)^{-1} [(\alpha+k\lambda)/\alpha]^{-1/k}.$$

Let α_4 be the α value so that the probability in (B.3) is maximized for fixed V_0 , i.e., α_4 satisfies

$$(B.5) \quad c(\alpha_4) = \sup_{\alpha} c(\alpha).$$

Theorem B. For every n and every k , we have

$$(B.6) \quad \alpha_4 = \frac{1}{2} [\sqrt{(k\lambda)^2 + 2k\lambda(2-\lambda) + \lambda^2} - (k-1)\lambda].$$

In the special case if $\lambda = 1$, (B.6) reduces to

$$(B.7) \quad \alpha_4 = 1.$$

Proof: Since

$$\frac{d}{d\alpha} \log c(\alpha) = \frac{-[\alpha^2 + (k-1)\lambda\alpha - k\lambda]}{\alpha(\alpha+k)(\alpha+k\lambda)} \begin{matrix} > \\ < \end{matrix} 0 \quad \text{for } \alpha \begin{matrix} < \\ > \end{matrix} \alpha_4,$$

$c(\alpha)$ achieves its maximum at α_4 and the theorem is proved.

Theorem B asserts that if all the population variances are equal, then sampling with equal number of observations is optimal for the construction of elliptical confidence regions. In this case there is no advantage of taking more observations from the control as we have had in the case of rectangular confidence regions when n is large.

References

- [1] Bechhofer, R. E. (1969). "Optimal allocation of observations when comparing several treatments with a control." International Symposium on Multivariate Analysis, 2, 463-473, Academic Press.
- [2] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press.
- [3] Dunnett, C. W. (1955). "A multiple comparison procedure for comparing several treatments with a control." J. Amer. Statist. Assoc., 50, 1096-1121.
- [4] Gupta, S. S. (1963). "Probability integrals of multivariate normal and multivariate t." Ann. Math. Statist., 34, 792-828.
- [5] Gupta, S. S. and Sobel, M. (1958). "On selecting a subset which contains all populations better than a standard." Ann. Math. Statist., 29, 235-244.
- [6] Lehmann, E. L. (1961). "Some model I problems of selection." Ann. Math. Statist., 32, 990-1012.
- [7] Robbins, H., Simons, G. and Starr, N. (1967). "A sequential analogue of the Behrens-Fisher problem." Ann. Math. Statist., 38, 1384-1391.
- [8] Robbins, H., Sobel, M. and Starr, N. (1968). "A sequential procedure for selecting the largest of k means." Ann. Math. Statist., 39, 88-92.
- [9] Šidák, Z. (1968). "On multivariate normal probabilities of rectangles." Ann. Math. Statist., 39, 1425-1434.
- [10] Tong, Y. L. (1969). "On partitioning a set of normal populations by their locations with respect to a control." Ann. Math. Statist., 40, 1300-1324.